

General Saddlepoint Approximations of Marginal Densities and Tail Probabilities

Riccardo Gatto and Elvezio Ronchetti[†]

September 1994

Revised: September 1995

Abstract

Saddlepoint approximations of marginal densities and tail probabilities of general nonlinear statistics are derived. They are based on the expansion of the statistic up to the second order. Their accuracy is shown in a variety of examples, including logit and probit models and rank estimators for regression.

Key Words: *bootstrap, logit model, marginal distribution, M-estimator, probit model, rank estimator, U-statistic, von Mises expansion.*

[†] Riccardo Gatto is Visitor, Department of Statistics, Stanford University, CA 94305-4065, USA; and Elvezio Ronchetti is Professor of Statistics, Faculty of Economic and Social Sciences, University of Geneva, CH-1211 Geneva, Switzerland. The authors thank A. C. Davison, T. J. DiCiccio, C. A. Field, S. Wang, the referees, the associate editor, and the editor for their helpful comments. The work of R. Gatto was partially supported by the Swiss National Science Foundation.

1 INTRODUCTION

Saddlepoint techniques are useful tools to derive very accurate approximations of densities and tail probabilities. Since their introduction into statistics in a seminal paper by H. E. Daniels (1954), they have been used successfully in a variety of situations. For a general review, see Reid (1988), Barndorff-Nielsen and Cox (1989), chap. 6, and Field and Ronchetti (1990). More recent applications can be found e.g. in Spady (1991), Butler, Huzurbazar and Booth(1992), Wood, Booth and Butler (1993), Wang (1993), and Ronchetti and Welsh (1994).

In this paper we focus on saddlepoint approximations of the marginal density and tail probabilities of a general statistic. Consider n i.i.d. observations Z_1, \dots, Z_n with underlying distribution F and a (possibly multivariate) statistic $T_n(Z_1, \dots, Z_n)$. Denote by $t_0 = T(F)$ the statistical functional defined by $T_n = T(F^{(n)})$, where $F^{(n)}$ is the empirical distribution function. Suppose we are interested in the density and tail probabilities of a real-valued function $m(T_n)$ with continuous and non-zero gradient at $t = t_0$, and continuous second derivative at $t = t_0$. This can be e.g. just a component of T_n or a linear contrast.

Our basic approximation can be derived as follows.

Step 1 Expand the statistic T_n up to the second order term

$$T_n - t_0 = \frac{1}{n} \sum_{i=1}^n k_1(Z_i; F) + \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n k_2(Z_i, Z_j; F) + O_p(n^{-\frac{3}{2}}). \quad (1)$$

Step 2 Compute the approximation of the variance of $m(T_n)$

$$\sigma_n^2 = \frac{1}{n} 4\sigma_g^2 + \frac{2}{n(n-1)} E \gamma^2(Z_1, Z_2) \quad (2)$$

and the approximation of the (standardized) cumulants $\kappa_{3n} = \kappa_3 n^{-\frac{1}{2}}$ and $\kappa_{4n} = \kappa_4 n^{-1}$, where κ_3 and κ_4 are $O(1)$ quantities given by

$$\kappa_3 = \sigma_g^{-3} \{E g^3(Z_1) + 3E [g(Z_1)g(Z_2)\gamma(Z_1, Z_2)]\}$$

$$\begin{aligned}\kappa_4 = & \sigma_g^{-4} \{ E g^4(Z_1) - 3\sigma_g^4 + 12E [g^2(Z_1)g(Z_2)\gamma(Z_1, Z_2)] \\ & + 12E [g(Z_1)g(Z_2)\gamma(Z_1, Z_3)\gamma(Z_2, Z_3)] \}. \end{aligned} \quad (3)$$

In these equations,

$$g(z) = \frac{1}{2}k_1^T(z; F) \frac{\partial}{\partial t} m(t) \big|_{t=t_0}, \quad (4)$$

$$\begin{aligned}\gamma(z_1, z_2) = & \frac{1}{2} \left(k_2^T(z_1, z_2; F) \frac{\partial}{\partial t} m(t) \big|_{t=t_0} \right. \\ & \left. + k_1^T(z_1; F) \frac{\partial^2}{\partial t \partial t^T} m(t) \big|_{t=t_0} k_1(z_2; F) \right), \end{aligned} \quad (5)$$

and

$$\sigma_g^2 = E g^2(Z_1), \quad (6)$$

and all the expectations are taken with respect to F . Define

$$\tilde{R}_n(\lambda) = \frac{1}{2}n\sigma_n^2\lambda^2 + \frac{1}{6}n^2\kappa_{3n}\sigma_n^3\lambda^3 + \frac{1}{24}n^3\kappa_{4n}\sigma_n^4\lambda^4. \quad (7)$$

Step 3 The saddlepoint approximations to the density and tail areas of $m(T_n) - m(t_0)$ are derived by applying the method of Easton and Ronchetti (1986) and are given by

$$g_n(v) = \left[\frac{n}{2\pi\tilde{R}_n''(\alpha)} \right]^{\frac{1}{2}} \exp\{n[\tilde{R}_n(\alpha) - \alpha v]\} \quad (8)$$

$$G_n(v) = P(m(T_n) - m(t_0) > v) = 1 - \Phi(r) + \phi(r) \left\{ \frac{1}{s} - \frac{1}{r} \right\} \quad (9)$$

where $s = \alpha[n\tilde{R}_n''(\alpha)]^{\frac{1}{2}}$, $r = \text{sgn}(\alpha)\{2n[\alpha v - \tilde{R}_n(\alpha)]\}^{\frac{1}{2}}$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution functions of the standard normal and $\alpha = \alpha(v)$ is the saddlepoint defined as the solution of the equation

$$\tilde{R}_n'(\alpha) = v. \quad (10)$$

The relative error of the approximation given in Step 3 is of order $O(n^{-1})$ and improves upon the standard normal which has error $O(n^{-\frac{1}{2}})$. Moreover, the

renormalization of the right hand side of (8) to make $\int g_n(v)dv = 1$ decreases the relative error to $O(n^{-\frac{3}{2}})$ in the normal region. Note that we do not assume any specific underlying parametric model and we require only that the statistic T_n admits the expansion given in Step 1. The approximation will be derived in Section 2.

There are three key ingredients in our procedure. First of all an expansion of the statistic up to the second order is needed. In the univariate case, Davison and Hinkley (1988) and Wang (1990b) have stressed the need to go beyond the linear term in the expansion of a nonlinear statistic in order to keep the order $O(n^{-1})$ for the relative error of the saddlepoint approximation. However, the approximation proposed by Davison and Hinkley (1988) does not seem to lead to accurate results whereas Wang (1990b) computes the saddlepoint approximation only in the special case where $k_2(Z_1, Z_2; F) = c \cdot q(Z_1; F)q(Z_2; F)$, for some function q and some constant c . In this paper we do not assume any special form for the kernels k_1 and k_2 and we consider the case of a multivariate statistic T_n . The computation of these kernels is not particularly complicated and will be discussed in Section 2. Moreover, in the Appendix we give the explicit forms of k_1 and k_2 for the classes of M -estimators and R -estimators.

The second ingredient is the approximation of $m(T_n) - m(t_0)$ up to the second order term by Taylor expansion from the corresponding expansion of $T_n - t_0$ obtained in Step 1,

$$\begin{aligned}
m(T_n) - m(t_0) &= (T_n - t_0)^T \frac{\partial}{\partial t} m(t) \big|_{t=t_0} \\
&\quad + \frac{1}{2} (T_n - t_0)^T \frac{\partial^2}{\partial t \partial t^T} m(t) \big|_{t=t_0} (T_n - t_0) + O_p(n^{-\frac{3}{2}}) \\
&= \frac{1}{n} \sum_{i=1}^n \left[k_1^T(Z_i; F) \frac{\partial}{\partial t} m(t) \big|_{t=t_0} \right] \\
&\quad + \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \left[k_2^T(Z_i, Z_j; F) \frac{\partial}{\partial t} m(t) \big|_{t=t_0} \right]
\end{aligned}$$

$$+k_1^T(Z_i; F) \frac{\partial^2}{\partial t \partial t^T} m(t) |_{t=t_0} k_1(Z_j; F) \Big] + O_p(n^{-\frac{3}{2}}) \quad (11)$$

This expression can be viewed as a U -statistic of degree 2; see Serfling (1980). Therefore, we can use the available expressions for the cumulants given in Step 2. Note that one can consider a nonparametric version of the approximation by replacing the expectations with respect to F with expectations under the empirical distribution function $F^{(n)}$, i.e. by replacing all the expectations with averages over the data.

The final ingredient in Step 3 is the general saddlepoint approximation of the density and tail area of a univariate statistic when approximations for the first four cumulants are available; see Easton and Ronchetti (1986) and Wang (1992).

The paper is organized as follows. In Section 2 we derive our basic approximation and we discuss some related approximations. In Section 3 we apply our approximation to a variety of models and estimators. They include, the maximum likelihood estimator for the parameters in regular logit and probit models, and the R -estimators in linear models. Small simulation studies and examples based on real data will show the accuracy and the generality of our approximation. Finally, in Appendixes A and B, we provide explicit formulas for the kernels k_1 and k_2 of multivariate M -estimators and R -estimators for regression models.

2 DERIVATION OF MARGINAL APPROXIMATIONS

As mentioned in the introduction, the first step is the expansion of the statistic up to the second order term. Let T_n be a multivariate statistical functional $T_n = T(F^{(n)})$, where $F^{(n)}$ is the empirical distribution function and let $t_0 = T(F)$. Then, a von Mises expansion (von Mises, 1947, Fernholz, 1984) of $T(F^{(n)}) - T(F)$, leads to (1), where the remainder is $O_p(\|F_n - F\|^3) = O_p(n^{-\frac{3}{2}})$. Because the error term in (1) is of order $n^{-\frac{3}{2}}$, the use of the quadratic approximation will produce a relative error of order n^{-1} in calculating the distribution of $m(T_n)$. The kernels (or differentials) k_1 and k_2 have the properties $\int k_1(z; F) dF(z) = 0$, $\int k_2(z_1, z, F) dF(z) = 0$, and $k_2(z_1, z_2; F) = k_2(z_2, z_1; F)$, $\forall z_1, z_2$. The first order kernel k_1 is usually called influence function of T and plays an important role in robust statistic; see Hampel (1974) and Hampel, Ronchetti, Rousseeuw, Stahel (1986). The kernels of a von Mises expansion can be computed iteratively as follows.

Theorem (Withers, 1983) Let $F_{\lambda, z} = (1 - \lambda)F + \lambda\Delta_z$, where Δ_z is the Dirac's distribution function which assigns mass one at z and assume that the $(i + 1)$ th differential of $T(\cdot)$ at F exists. Then

$$k_1(z; F) = \frac{\partial}{\partial \lambda} T(F_{\lambda, z})|_{\lambda=0} \quad (12)$$

and

$$k_{i+1}(z_1, \dots, z_{i+1}; F) = \frac{\partial}{\partial \lambda} k_i(z_1, \dots, z_i; F_{\lambda, z_{i+1}})|_{\lambda=0} + \sum_{r=1}^i k_i\langle z_1, \dots, z_{i+1}; F \rangle_r, \quad (13)$$

where $\langle \cdot \rangle_r$ indicates that the r th argument has been dropped.

In our case we can apply (13) with $i = 1$ and obtain

$$k_2(z_1, z_2; F) = \frac{\partial}{\partial \lambda} k_1(z_1; F_{\lambda, z_2})|_{\lambda=0} + k_1(z_2; F). \quad (14)$$

In Appendixes A and B we give explicit formulas for k_1 and k_2 in the case of multivariate M -estimators and R -estimators for regression models.

Let us now consider the statistic $m(T_n) - m(t_0)$. By a Taylor expansion at t_0 and by using (1), we can obtain the following quadratic approximation to this statistic

$$U_n = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n h(Z_i, Z_j), \quad (15)$$

where $h(z_1, z_2) = \left\{ \left(k_1^T(z_1; F) + k_1^T(z_2; F) + k_2^T(z_1, z_2; F) \right) \partial m(t) / \partial t \big|_{t=t_0} + k_1^T(z_1; F) \partial^2 m(t) / \partial t \partial t^T \big|_{t=t_0} k_1(z_2; F) \right\} / 2$. U_n is a U -statistic of degree 2, see Serfling (1980), chap. 5. This statistic admits an Edgeworth expansion and the cumulants can be obtained by using formulas (2) and (3), see Bickel, Götze, van Zwet (1986). It can be shown that, if σ_n^{*2} , $\kappa_{3n}^* \sigma_n^{*3}$ and $\kappa_{4n}^* \sigma_n^{*4}$ denote the exact second, third and fourth cumulants of $m(T_n) - m(t_0)$, then we have the following relations: $\sigma_n^{*2} = \sigma_n^2 + O(n^{-2})$, $\kappa_{3n}^* \sigma_n^{*3} = \kappa_{3n} \sigma_n^3 + O(n^{-3})$ and $\kappa_{4n}^* \sigma_n^{*4} = \kappa_{4n} \sigma_n^4 + O(n^{-3})$. These errors will not increase the error in the saddlepoint approximation. Hence, truncating the von Mises expansion at the second order term will change these cumulants with sufficiently small errors and, this will result in a relative error $O(n^{-1})$ in the saddlepoint approximation. U_n is a univariate statistic for which we know approximations of the first four cumulants to the appropriate order. Therefore, we can apply the general saddlepoint approximation derived by Easton and Ronchetti (1986). This leads to the approximations (8) and (9).

It was pointed out by Wang (1992) that (10) could have multiple solutions and in this case the approximation could fail in some areas of the distribution. To avoid this problem, he proposed a modification of $\tilde{R}_n(\lambda)$ of the form

$$R_n^W(\lambda; b) = \frac{1}{2} n \sigma_n^2 \lambda^2 + \left(\frac{1}{6} n^2 \kappa_{3n} \sigma_n^3 \lambda^3 + \frac{1}{24} n^3 \kappa_{4n} \sigma_n^4 \lambda^4 \right) \omega_n(\lambda; b), \quad (16)$$

where $\omega_n(\lambda; b) = \exp\{-n \sigma_n^2 b^2 \lambda^2 / 2\}$ and $b = \max\{1/2, \inf\{a \mid R_n^{W''}(\lambda; a) > 0\}\}$. Replacing $\tilde{R}_n(\lambda)$ by $R_n^W(\lambda; b)$ in (8), (9) and (10) guarantees a unique solution of

(10) and does not change the order of the approximations. For details, we refer to Wang (1992).

To ensure the validity of our approximation we need two sets of conditions. The first one is needed to ensure the validity of the approximation of the statistic T_n up to the quadratic term in Step 1. A discussion of these assumptions can be found for example in Serfling (1980), sec. 6.2, pp. 214-221. They hold for instance for M - and R -estimators with smooth score functions but they are not satisfied for sample quantiles. The second set of conditions is given in Bickel, Götze and van Zwet (1986), p. 1465, or Field and Ronchetti (1990) Th. 2.4, pp. 15-16. They guarantee the correct orders of the cumulants in Step 2 and the validity of the Edgeworth expansion of the corresponding U -statistic up to $O(n^{-\frac{3}{2}})$.

There has been some recent work to derive saddlepoint approximations of marginal densities and tail probabilities. Jing and Robinson (1994) provide approximations to marginal probabilities for statistics which can be expressed as smooth transformations of multivariate means. The approximations are obtained by tilting the Edgeworth expansions available for this case. The conditions are those required for the validity of such Edgeworth expansions; see Robinson, Höglund, Holst and Quine (1990). Our technique seems to be more general and can be applied to statistics which are not necessarily smooth functions of multivariate means as for instance, M -estimators and rank estimators. Also, since we approximate directly the marginal statistic $m(T_n) - m(t_0)$, our saddlepoint equation (10) is univariate and easy to solve numerically. Jing and Robinson's technique requires solving a saddlepoint equation of dimension $\dim(T_n) - 1$ (cf. Jing and Robinson, 1994, bottom of p. 1121) and this is generally harder especially when $\dim(T_n)$ is moderate to large.

Fan and Field (1995) derive saddlepoint approximations to marginal densities

for the class of M -estimators. They start with the saddlepoint of the joint density of an M -estimator given by Field (1982) and compute the marginal by means of the formula given in DiCiccio, Field and Fraser (1990) and DiCiccio and Martin (1991). Our approximation does not require an approximation of the joint density but our expansion of $T_n - t_0$ in Step 1 could in principle be used to obtain such an approximation. Then, the formula derived in the papers mentioned above can be applied to obtain an approximation to the marginal density and tail area. A saddlepoint approximation of the distribution of a univariate smooth function of an M -estimator in the nonparametric situation is derived also by Davison, Hinkley and Worton (1995). Their technique uses an approximation to the joint density of M -estimators and Laplace's method to approximate the integrals and obtain the marginal distribution.

Finally, a related paper is Chen and Do (1994) where the authors approximate a statistic T_n which can be expressed as a smooth function of multivariate means up to the quadratic term. Through a formula given in Barndorff-Nielsen and Cox (1989), p. 77, they then approximate the distribution function of such statistics in terms of the cumulative distribution function and the density function of the linear term, and the conditional expectation of the quadratic term given the linear term. They propose to approximate the cumulative distribution and density of the linear term by means of a saddlepoint approximation and to approximate the conditional expectation by Monte Carlo techniques. While this is an interesting device to reduce the computational burden of a direct bootstrap procedure on the statistic T_n , the technique still requires sophisticated resampling in order to approximate the conditional expectation. This seems generally difficult to obtain. Our approximation is more direct, automatic, and less computationally intensive.

Note that we can derive a nonparametric version of the approximation by replacing in formulas (2) to (5) the expectations with respect to F by the ex-

pectations with respect to the empirical distribution $F^{(n)}$. Wang (1990a, 1992) studied the nonparametric version of the general saddlepoint approximation used above to derive (8) and (9) and compared it to the bootstrap.

3 APPLICATIONS

In this section we present a variety of applications which show the accuracy and the generality of our approximation. The computations were carried out with Matlab software on a Sun workstation. Pseudo-random numbers were generated by the function “rand”, available in Matlab. This function is based on a linear congruential method.

3.1 Logit Model

Consider a response variable y with values 0 or 1, a p -dimensional vector x of explanatory variables, and define $P(y = 1 \mid x^T \beta) = \pi(x^T \beta) = 1/(1 + \exp\{-x^T \beta\})$, where $\beta \in \mathbb{R}^p$. The maximum likelihood estimator (MLE) for β is the solution of the system of equations $\sum_{i=1}^n s(x_i, y_i; \beta) = 0$, where the score function s is given by $s(x, y; \beta) = [y - \pi(x^T \beta)]x$. Therefore, the MLE for β is an M -estimator with ψ -function equal to the score function. We can then apply our approximation by using the kernels reported in Appendix A.

We performed a small simulation study for a logit model with a constant term and two regression coefficients for sample size $n = 20$. The x 's are redrawn for each sample and are generated by taking the exponential of uniform drawings in $[0, 1]$ (x_{1i}) and the square of uniform drawings in $[0, 1]$ (x_{2i}). The true values used in the simulations are $(\beta_0, \beta_1, \beta_2) = (2, -1, -2)$. The exact results are obtained by means of 50000 Monte Carlo simulations. We compare the nonparametric version of our approximation, the bootstrap approximation, and the asymptotic normal approximation. The saddlepoint approximation plotted in the graphics corresponds to an average of 1000 simulated saddlepoint approximations. Each simulated curve is recentered and we use Wang's modification given by (16). The asymptotic normal distribution is obtained via the estimation of the asymptotic

variance by

$$V(\beta, F^{(n)}) = \frac{1}{n} M^{-1}(s, F^{(n)}) Q(s, F^{(n)}) [M^{-1}(s, F^{(n)})]^T, \quad (17)$$

where $F^{(n)}$ is the empirical distribution of the observations and $M(s, F^{(n)}) = -n^{-1} \sum_{i=1}^n \partial s(x_i, y_i; \beta) / \partial \beta |_{\beta=\hat{\beta}}$, $Q(s, F^{(n)}) = n^{-1} \sum_{i=1}^n s(x_i, y_i; \hat{\beta}) s^T(x_i, y_i; \hat{\beta})$, $\hat{\beta}$ being the maximum likelihood estimator. We found that in this case and with small sample sizes, the “sandwich estimator” of the asymptotic variance given by (17) turns out to be more reliable than the asymptotically equivalent estimators $n^{-1} M^{-1}(s, F^{(n)})$ or $n^{-1} Q^{-1}(s, F^{(n)})$. The asymptotic normal distribution is an average of 1000 simulations. The bootstrap distribution is based on 1000 simulated samples. For each of these samples, the bootstrap distribution is estimated by means of 100 resamplings.

Figure 1 compares the saddlepoint and the asymptotic normal approximations with the “exact” (histogram based on 50000 simulations), for the marginal density of the constant term. Figure 2 gives the corresponding Q-Q plot of the distribution of the constant term and Figure 3 the right tail area. The nonparametric saddlepoint approximations are very accurate, especially in the tails. The 95% confidence intervals given in Figure 3 show that the variability of the nonparametric saddlepoint and bootstrap approximations is small and the differences are clearly significant. The standard errors of the tail areas between 0.900 and 0.975 are always smaller than $2 \cdot 10^{-3}$. Figure 4 compares the bootstrap density based on 100 resamplings with the exact density obtained by smoothing the histogram based on 50000 simulations. It is clear that this approximation is not accurate, especially in the left tail, and one would need many more resamplings in order to attain the accuracy of the nonparametric saddlepoint approximation. Table 1 shows the probabilities in the left tail of the densities of the second coefficient, and Table 2 refers to the right tail of the sum of the constant term with the first coefficient. In these tables, the probability P^e is obtained from the

Figure 1: *Nonparametric saddlepoint and asymptotic normal approximations of the exact marginal density of $\hat{\beta}_0$ (histogram based on 50000 simulations) in the logit model*

Figure 2: *Q-Q plot of the quantiles of the nonparametric saddlepoint, asymptotic normal and bootstrap approximations with respect to the quantiles of the exact cumulative distribution of $\hat{\beta}_0$ in the logit model. The solid straight line is the identity, corresponding to the quantiles of the exact distribution.*

simulated exact distribution, P^s is obtained from the saddlepoint density by using a simple univariate trapezoidal rule and P^a is computed from the asymptotic normal distribution.

Similar results (not shown here) are obtained for the first coefficient and for $n = 16$; cf. Gatto (1994).

Figure 3: *Nonparametric saddlepoint, asymptotic normal and bootstrap approximations of the exact right tail of $\hat{\beta}_0$ in the logit model. 95% confidence intervals for the nonparametric saddlepoint and bootstrap approximations for the tail areas corresponding to 0.900, 0.925, 0.950, and 0.975 are indicated.*

Figure 4: *Bootstrap approximation based on 100 resamplings of the exact marginal density of $\hat{\beta}_0$ in the logit model*

3.2 Probit Model

We have the same model as in section 3.1 with $P(y = 1 \mid x^T\beta) = \pi(x^T\beta) = \Phi(x^T\beta)$ and $s(x, y; \beta) = [y - \Phi(x^T\beta)]x\phi(x^T\beta)/\{\Phi(x^T\beta)[1 - \Phi(x^T\beta)]\}$ where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and the cumulative of the standard normal distribution.

We consider the following data set from Press and Wilson (1978). Population change data were collected for some states of the U.S. in an effort to explain population changes from the 1960 census to the 1970 census. Although the original data set has 50 observation and five explanatory variables, we consider only 24 observations, which correspond to the coast states and two explanatory variables: the income (x_1) and the number of births (x_2). The dependent variable (y) equals 0 or 1, according to whether the change is below or above the median change for all states. In this nonparametric case we investigate the performance of our saddlepoint approximation and normal asymptotic distribution as approximations to the bootstrap distribution based on 10000 resamplings (histogram in Figure 5). It should be noticed that this is not necessarily close to the true unknown distribution of the estimators. We increased the number of bootstrap replicates but this does not affect the conclusions of our comparison. All densities are recentered.

Figure 5 shows the approximations to the marginal density of $\hat{\beta}_2$. It is clear that the saddlepoint approximation improves upon the asymptotic normal approximation and is close to the bootstrap but avoids the resampling.

3.3 R -estimators

In order to overcome some of the deficiencies of least-squares estimators (lost of efficiency in the presence of small departures from the normality assumption on the errors, sensitivity to outliers, etc.), one can use regression R -estimators (rank estimators). Basically, in the estimating equation of an R -estimator the ranks of

Figure 5: *Nonparametric saddlepoint and asymptotic normal approximations of the bootstrap marginal density of $\hat{\beta}_2$ (histogram based on 10000 resamplings) in the probit model for the data set of Press and Wilson (1978)*

the regression residuals appear instead of the residuals themselves. Regression R -estimators inherit robustness properties from the rank tests from which they are derived and are less affected by small departures from normality assumption on the error distribution. A basic reference is Hettmansperger (1984). More recent results can be found in Naranjo and Hettmansperger (1994).

Consider the linear regression model

$$y = u\gamma + X\beta + \varepsilon, \quad (18)$$

where y is a $n \times 1$ vector, u denotes a $n \times 1$ vector of “ones”, X is a $n \times p$ matrix of explanatory variables, $\gamma \in \mathbb{R}$ and $\beta \in \mathbb{R}^p$ are unknown parameters, and ε is a $n \times 1$ vector of i.i.d. random errors with mean zero. Define a non-constant sequence of scores $a(1) \leq a(2) \leq \dots \leq a(n)$ such as $a(k) + a(n - k + 1) = 0$ and $\sum_{i=1}^n a(i) = 0$. Then, an R -estimator of the regression coefficients β is the value $\hat{\beta}$ which minimizes $\sum_{i=1}^n a[R(y_i - x_i^T \beta)](y_i - x_i^T \beta)$, where x_i is the i th row of X and $R(y_i - x_i^T \beta)$ is the rank of $y_i - x_i^T \beta$ among $y_1 - x_1^T \beta, \dots, y_n - x_n^T \beta$. Taking the derivative with respect to β (which exist almost everywhere) gives the estimating

equation

$$\sum_{i=1}^n a[R(y_i - x_i^T \beta)] x_i \simeq 0. \quad (19)$$

The scores are typically generated by a continuous and non-decreasing function $J : [0, 1] \rightarrow \mathbb{R}$ which satisfies the properties $\int_0^1 J(v) dv = 0$ and $\int_0^1 J^2(v) dv = 1$. For instance, the Wilcoxon scores are obtained by $a(i) = J(i/(n+1))$, where $J(u) = \sqrt{12}(u - 1/2)$.

We compare the performance of our approximation on a data set from Becker and Harmett (1987). The data concern 20 cities and represent the sales of hospital supplies (y) for a given firm, the cost of advertising (x_1), the number of hospitals in each city (x_2) and the the number of competing firms in each city (x_3). We consider the linear model (18). The values of the coefficients of the R -estimator are $\hat{\beta}_1 = 1.81$, $\hat{\beta}_2 = 25.82$ and $\hat{\beta}_3 = -42.16$.

We compare the marginal density of the second estimated coefficient for the bootstrap distribution based on 10000 resamplings, the asymptotic normal distribution by using the estimated asymptotic variance, and the nonparametric saddlepoint approximation. As we mention in Appendix B, the computation of the estimated asymptotic normal distribution and the nonparametric saddlepoint approximation require the estimation of the conditional density of $y_i \mid x_i^T \beta$ from the residuals. We use a kernel density estimate with the bandwidth determined by means of cross-validation. Figure 6 shows once again the accuracy of the saddlepoint approximation which picks up the skewness of the density.

Figure 6: *Nonparametric saddlepoint and asymptotic normal approximations of the bootstrap marginal left tail of $\hat{\beta}_2$ in the regression model for the data set of Becker and Harmett (1987). The scale on the right axis is P and on the left axis is $\log\{P/(1 - P)\}$.*

Table 1: *Nonparametric saddlepoint and asymptotic normal approximations of the exact tail probabilities of $\hat{\beta}_2$ in the logit model*

t	$P^e(\hat{\beta}_2 < t)$	$P^s(\hat{\beta}_2 < t)$	$P^a(\hat{\beta}_2 < t)$
-10.44	0.010	0.009	0.002
-9.36	0.015	0.012	0.003
-8.28	0.020	0.016	0.005
-7.20	0.028	0.023	0.007
-6.12	0.039	0.033	0.012
-5.76	0.044	0.037	0.014
-5.40	0.049	0.043	0.018
-5.04	0.056	0.049	0.022
-4.68	0.064	0.057	0.028
-4.32	0.072	0.070	0.035
-3.96	0.083	0.079	0.045
-3.60	0.095	0.093	0.057
-3.24	0.110	0.111	0.074
-2.88	0.126	0.132	0.096
-2.52	0.146	0.158	0.124
-2.16	0.170	0.190	0.159
-1.80	0.199	0.227	0.202
-1.44	0.231	0.271	0.253
-1.08	0.267	0.321	0.311
-0.72	0.314	0.375	0.376

Table 2: *Nonparametric saddlepoint and asymptotic normal approximations of the exact tail probabilities of $\hat{\beta}_0 + \hat{\beta}_1$ in the logit model*

t	$P^e(\hat{\beta}_0 + \hat{\beta}_1 < t)$	$P^s(\hat{\beta}_0 + \hat{\beta}_1 < t)$	$P^a(\hat{\beta}_0 + \hat{\beta}_1 < t)$
0.26	0.653	0.642	0.626
0.52	0.718	0.709	0.673
0.78	0.772	0.768	0.757
1.04	0.817	0.818	0.826
1.30	0.852	0.858	0.879
1.56	0.880	0.889	0.900
1.82	0.903	0.913	0.932
2.08	0.921	0.931	0.954
2.34	0.935	0.944	0.969
2.60	0.945	0.955	0.974
2.86	0.954	0.963	0.982
3.12	0.961	0.969	0.987
3.38	0.967	0.974	0.991
3.64	0.971	0.978	0.992
3.90	0.975	0.981	0.994
4.16	0.978	0.984	0.995
4.42	0.981	0.986	0.996
4.68	0.983	0.988	0.996
4.94	0.985	0.989	0.997
5.20	0.987	0.991	0.998

4 CONCLUSION

In this paper we present a simple and accurate approximation to marginal distributions which can be applied to fairly general problems. A small simulation and examples based on real data sets show the accuracy of the method. In the nonparametric case, this approximation can be viewed as an alternative to the Monte Carlo resampling. The accuracy of the saddlepoint approximation is substantially greater than that of the normal approximation, while the computing time is much shorter than that required by Monte Carlo resampling.

We mentioned in Section 2 that alternative approximations have been proposed very recently for specific situations. More research is needed to establish the exact relationship among them.

APPENDIX A: KERNELS k_1 AND k_2 IN THE VON MISES EXPANSION OF MULTIVARIATE M -ESTIMATORS

Given a sample Z_1, \dots, Z_n of i.i.d. observations with common distribution F , consider a p -dimensional M -estimator T_n , i.e. the solution of the system of equations $\sum_{i=1}^n \psi(Z_i, T_n) = 0$, for some function $\psi(\cdot, \cdot)$. Let $M(\psi, F) = - \int \partial \psi(z, t) / \partial t |_{t=t_0} dF(z)$, then it is well-known that the influence function of the M -estimator is given by

$$k_1(z, F) = M^{-1}(\psi, F) \psi(z, t_0),$$

where $t_0 = T(F)$ is the solution of $\int \psi(z, T(F)) dF(z) = 0$; cf. Hampel, Ronchetti, Rousseeuw and Stahel (1986). With some algebra, an application of (14) gives the second order kernel

$$\begin{aligned} k_2(z_1, z_2, F) = & \\ & k_1(z_1, F) + k_1(z_2, F) + M^{-1}(\psi, F) \left\{ v(z_1, z_2, F) \right. \\ & \left. + \frac{\partial}{\partial t} \psi(z_1, t) |_{t=t_0} k_1(z_2, F) + \frac{\partial}{\partial t} \psi(z_2, t) |_{t=t_0} k_1(z_1, F) \right\}, \end{aligned}$$

with

$$\begin{aligned} v(z_1, z_2, F) = & \left(k_1^T(z_2, F) \int \frac{\partial^2}{\partial t \partial t^T} \psi_1(z, t) |_{t=t_0} dF(z) k_1(z_1, F), \right. \\ & \left. \dots, k_1^T(z_2, F) \int \frac{\partial^2}{\partial t \partial t^T} \psi_p(z, t) |_{t=t_0} dF(z) k_1(z_1, F) \right)^T \end{aligned}$$

APPENDIX B: KERNELS k_1 AND k_2 IN THE VON MISES EXPANSION OF R -ESTIMATORS FOR REGRESSION

The estimating equation of an R -estimator is given by (19). The functional form of a regression R -estimator $T(H)$ is implicitly defined by

$$\int_{x,y} J[F(y - x^T T(H))] x dH(x, y) = 0,$$

where F and f are the distribution and density of the random variable y conditional on x , and H is the joint distribution of (x_i^T, y_i) , with x_i the i th row of the design matrix X .

According to this functional expression, the first order kernel (influence function) is given by

$$k_1(x, y; F, H) = J[F(y - x^T T(H))] M^{-1}(J, F, H) x,$$

where $M(J, F, H) = \int_{x,y} x x^T J'[F(y - x^T T(H))] f(y - x^T T(H)) dH(x, y)$ and J' is the derivative of J ; see Hettmansperger (1984) for details.

The second order kernel can be computed by means of (14),

$$\begin{aligned} k_2(x_1, y_1, x_2, y_2, F, H) = & \\ & k_1(x_1, y_1, F, H) + k_1(x_2, y_2, F, H) + M^{-1}(J, F, H) \\ & \left\{ \int_{x,y} x (J''[F(y - x^T T(H))] f^2(y - x^T T(H)) + J'[F(y - x^T T(H))] \right. \\ & f'(y - x^T T(H))) x^T k_1(x_1, y_1, F, H) x^T k_1(x_2, y_2, F, H) dH(x, y) \\ & - x_2 x_2^T J'[F(y_2 - x_2^T T(H))] f(y_2 - x_2^T T(H)) k_1(x_1, y_1, F, H) \\ & \left. - x_1 x_1^T J'[F(y_1 - x_1^T T(H))] f(y_1 - x_1^T T(H)) k_1(x_2, y_2, F, H) \right\} \end{aligned}$$

Note that, in contrast with M -estimators, the influence function and asymptotic variance depend on the underlying conditional density f . In order to compute the nonparametric version of the asymptotic variance, it is then necessary

to estimate the conditional density f from the residuals. For this purpose we use kernel density estimation techniques.

REFERENCES

- Barndorff-Nielsen, O. E., Cox, D. R. (1989), *Asymptotic Techniques for Use in Statistics*, Chapman & Hall.
- Becker, W. E., Harmett, D. L. (1987), *Business and Economics Statistics with Computer Applications*, Addison-Wesley.
- Bickel, P. J., Götze, F., van Zwet, W. R. (1986), “The Edgeworth Expansion for U -Statistics of Degree Two”, *The Annals of Statistics*, 14, 1463-1484.
- Butler, R. W., Huzurbazar, S., Booth, J. G. (1992), “Saddlepoint Approximations for the Barlett-Nanda-Pillai Trace Statistic in Multivariate Analysis”, *Biometrika*, 79, 705-715.
- Chen, Z., Do, K. A. (1994), “The Bootstrap Method with Saddlepoint Approximations and Importance Sampling”, *Statistica Sinica*, 4, 407-421.
- Daniels, H. E. (1954), “Saddlepoint Approximations in Statistics”, *Annals of Mathematical Statistics*, 25, 631-650.
- Davison, A. C., Hinkley, D. V. (1988), “Saddlepoint Approximations in Resampling Methods”, *Biometrika*, 75, 417-431.
- Davison, A. C., Hinkley, D. V., Worton, B. J. (1995), “Accurate and Efficient Construction of Bootstrap Likelihoods”, Manuscript.
- DiCiccio, T. J., Field, C. A. and Fraser, D. A. S. (1990), “Approximations of Marginal Tail Probabilities and Inference for Scalar Parameters”, *Biometrika*, 77, 77-95.
- DiCiccio, T., Martin, M. A. (1991), “Approximations of Marginal Tail Probabilities for a Class of Smooth Functions with Applications to Bayesian and Conditional Inference”, *Biometrika*, 78, 891-902.

- Easton, G. S., Ronchetti, E. (1986), “General Saddlepoint Approximations with Applications to L Statistics”, *Journal of the American Statistical Association*, 81, 420-430.
- Fan, R. Y. K., Field, C. A. (1995), “Approximations for Marginal Densities of M -Estimators”, *The Canadian Journal of Statistics*, to appear.
- Fernholz, L. T. (1983), *Von Mises Calculus for Statistical Functionals*, Springer-Verlag.
- Field, C. (1982), “Small Sample Asymptotic Expansions for Multivariate M -Estimates”, *The Annals of Statistics*, 10, 672-689.
- Field, C. A., Ronchetti, E. (1990), *Small Sample Asymptotics*, IMS Monograph Series, 13, Hayward (CA).
- Gatto, R. (1994), *Saddlepoint Methods and Nonparametric Approximations for Econometric Models*, Ph.D. Dissertation, University of Geneva, Switzerland.
- Hampel, F. R. (1974), “The Influence Curve and its Role in Robust Estimation”, *Journal of the American Statistical Association*, 69, 383-393.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986), *Robust Statistics. The Approach Based on Influence Functions*, Wiley & Sons.
- Hettmansperger, T. P. (1984), *Statistical Inference Based on Ranks*, Wiley & Sons.
- Jing, B., Robinson, J. (1994), “Saddlepoint Approximations for Marginal and Conditional Probabilities of Transformed Variables”, *The Annals of Statistics*, 22, 1115-1132.

- Naranjo, J. D., Hettmansperger, T. P. (1994), “Bounded Influence Rank Regression”, *Journal of the Royal Statistical Society*, Ser. B, 56, 209-220.
- Press, S. J., Wilson, S. (1978), “Choosing Between Logistic Regression and Discriminant Analysis”, *Journal of the American Statistical Association*, 73, 699-705.
- Reid, N. (1988), “Saddlepoint Methods and Statistical Inference”, *Statistical Science*, 3, 213-238.
- Robinson, J., Höglund, T., Holst, L., Quine, M. P. (1990), “On Approximating Probabilities for Small and Large Deviations in \mathbb{R}^d ”, *Annals of Probability*, 18, 727-753.
- Ronchetti, E., Welsh, A. H. (1994), “Empirical Saddlepoint Approximations for Multivariate M -Estimators”, *Journal of the Royal Statistical Society*, Ser. B, 56, 313-326.
- Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley & Sons.
- Spady, R. H. (1991), “Saddlepoint Approximations for Regression Models”, *Biometrika*, 78, 879-889.
- Von Mises, R. (1947), “On the Asymptotic Distribution of Differentiable Statistical Functions”, *Annals of Mathematical Statistics*, 18, 309-348.
- Wang, S. (1990a), “Saddlepoint Approximations in Resampling Analysis”, *Annals of the Institute of Statistical Mathematics*, 42, 115-131.
- Wang, S. (1990b), “Saddlepoint Approximations for Nonlinear Statistics”, *The Canadian Journal of Statistics*, 18, 225-264.
- Wang, S. (1992), “General Saddlepoint Approximations in the Bootstrap”, *Statistics & Probability Letters*, 13, 61-66.

- Wang, S. (1993), “Saddlepoint Expansions in Finite Population Problems”, *Biometrika*, 80, 583-590.
- Withers, C. S. (1983), “Expansion for the Distribution and quantiles of a regular functional of the empirical distribution with applications to non-parametric confidence intervals”, *The Annals of Statistics*, 11, 577-587.
- Wood, A. T. A., Booth, J. G., Butler, R. W. (1993), “Saddlepoint approximation to the CDF of some statistics with nonnormal limit distribution”, *Journal of the American Statistical Association*, 88, 680-686.